RedFame

Journal of Education and Training Studies
Vol. 4, No. 7; July 2016
ISSN 2324-805X E-ISSN 2324-8068
Published by Redfame Publishing
URL: http://jets.redfame.com

Analysis of High School German Textbooks through Rasch Measurement Model

Veli Batdı[1], Şenel Elaldı[2]

[1]Muallim Rıfat Education Faculty, Department of Educational Sciences, University of 7 Aralık, Kilis, Turkey

[2]Foreign Languages Department, University of Cumhuriyet, Sivas, Turkey

Correspondence: Şenel Elaldı, Foreign Languages Department, University of Cumhuriyet, Sivas, Turkey

Received: March 14, 2016 Accepted: April 2, 2016 Online Published: April 7, 2016

doi:10.11114/jets.v4i7.1455 URL: http://dx.doi.org/10.11114/jets.v4i7.1455

Abstract

The purpose of the present study is to analyze German teacher trainers' views on high school German textbooks through the Rasch measurement model. A survey research design was employed and study group consisted of a total of 21 teacher trainers, three from each region and selected randomly from provinces which are located in seven regions and categorized as developed, moderately developed, and least developed. The study Data were collected through a questionnaire developed by the researchers in the light of experts' views. When content validity indices (CVIs) and content validity ratios (CVRs) of the questionnaire items were calculated, the result (CVI>CVR/0.82>0.56) indicated the questionnaire to be reliable. Three facets of the study according to the Rasch measurement model were Judges (21 German teacher trainers), items related to high school German textbooks (11 items) and German textbooks (A1.1, A1.2, A2.1, A2.2) for 9th, 10th, 11th, and 12th grades. According to the Rasch analysis results, while the textbook coded A1.1 has the highest quality, the textbook coded A2.2has the poorest quality. In terms of items, the most difficult item was 10 while the easiest item was 1, and for judges, J7 had the most severe while J9 had the most lenient behavior. In the light of the results, more rigorous and detailed studies are suggested to improve the quality of textbooks.

Keywords: German teacher trainer, German textbook, the Rasch measurement model, teachers' views

1. Introduction

Foreign language teaching has many important components but the essential constituents are materials that are used to increase learners' knowledge and/or experience of learning by many teachers. According to Tomlinson (2001, p. 66), materials "anything used to facilitate the learning of language" can be presented in print, through live performance or display, or on cassette, CD ROM, DVD or the Internet. Although grammar books and dictionaries were the only language teaching materials of past years used by teachers, today, there is a great variety of language teaching materials on the market (Crystal, 1987) containing visual, auditory, kinesthetic, studial, experiential, analytic, and global learning styles in themselves (Tomlinson, 2001).Therefore, today the scope of language learning materials includes not only purchased materials, but also materials that are provided online as well as those generated by the teacher and even the students (NCTE, 2014). In this sense, materials; (1) should be up to date (e.g. published within the past 10 years), (2) take into account the linguistic and cultural diversity of the student population, (3) conducive to being used with a variety of grouping strategies, and (4) contain exercises in which learners share previous experience with prior knowledge of the content (Wen-Cheng, Chien-Hung & Chung-Chieh, 2011)**.** Furthermore, when adapting and using materials which entails selecting appropriately, being creative, modifying, and supplementing in teaching and learning situations, (Dudley- Evans & St John, 1998), learners should be the center of instruction. However, in many cases, since teachers and students rely on materials, the materials become the center of instruction (Kitao & Kitao, 1997). On the other hand, while selection of the right materials makes teaching and learning a worthwhile activity and helps effective classroom environment, uninteresting and complicated materials lead learning to become a dull and monotonous activity (Dar, 2012). It is therefore necessary to select appropriate materials in order to adequately arise and maintain students' interest; at the same time they must be related positively to the aspects of their inner make up such as age, level of education, social attitudes, the intellectual ability and level of emotional maturity (Cunningsworth, 1995). In addition, materials should be at a slightly higher level of difficulty than the students' current level of foreign language proficiency (Kitao & Kitao, 1997). As explained by Küçükahmet (1995), other benefits of materials used in foreign

language classes are as follows : (a) provide economy in time and speech, (b) simplify the course, (c) make the course vivid and clear, (d) increase students' interest and motivation, (e) create desire of learning, (f) make abstract concepts concrete (g) enrich the course. Materials, therefore, should be selected and adapted carefully and the progress should be monitored to reveal whether they fulfill the needs of students.

Despite the variety of technology-based, innovative instructional materials in foreign language education these days, a textbook has always been the most preferred and basic tool in achieving aims and objectives concerning learner needs (Cunningsworth, 1995). Undeniably, "one of the most important decisions an instructor makes is the selection of a textbook" (Chatman & Goetz, 1985, p. 150). According to Williams (1983), ''the textbook is a tool and the teacher must know not only how to use it but how useful it can be'' (p. 254). In addition, textbooks enable teachers to organize their teaching (Richards & Renandya , 2002) by (1) assuring a measure of structure, consistency, and logical progression in a class, (2) minimizing preparation time for teachers, (3) providing novice teachers with guidance in course and activity design, and (4) providing multiple resources: tapes, CDs, videos, self-study workbooks etc. (Parrish, 2004).

Although using textbooks by sticking slavishly, from cover to cover, without any supplemental material is not preferred, both teachers and students need a framework on which to build and textbooks enable this (Garinger, 2002). In this context, many teachers use textbooks as 'bridges' to stimulate their thinking (Gray, 2002) ; resources rather than course materials used alone (Richards, 2001); and "structuring tools", providing convenient structure in teaching-learning system (Crawford, 2002, p.83). Furthermore, Sikorova (2011) identifies three approaches to textbook use as *adhering (or adopting)*, regarding them as the authority, *elaborating*, supplementing them with other resources, and *creating,* developing one's own units of study. However, considering textbooks as the authority without adaptations is a matter of debate. According to Nation and Macalister (2010), some reasons for doing adaptation are as follows: A textbook does not (1) include all the activities that the teacher has used successfully before, (2) contain content that is suitable for the learners' level of proficiency or age, (3) include language items, skills, ideas that the learners need.

According to Allwright (1982), although textbooks cannot cater for the needs in classrooms around the world, it is not recommend to be completely abandoned. In this regard teachers' role is not limited to transmit the content of printed materials, but their aim is to elicit "what students need to learn" (Sheikhzadeh Marand (2011; p. 553) and to select textbooks in line with students' needs. In parallel with this purpose, Cunningsworth (1995, p.7) states that "it is of crucial importance that careful selection is made, and that the materials selected closely reflect the needs of the learners and the aims, methods and values of the teaching program." Furthermore, without textbooks, a program may have no impact; therefore, they provide structure and a syllabus (Richards, 2001). From the foregoing, therefore, it is concluded that whether one believes that textbooks are too inflexible and biased to be used directly as instructional material, there can be no denying the fact that textbooks still maintain enormous popularity (Mohammadi & Abdi, 2014).

1.1 Purpose of the Research

Purpose of the present study is to analyze German teacher trainers' views on high school German textbooks through the Rasch measurement model. In line with this aim, the following sub-aims have been included in the study:

1. to perform a general analysis of views on high school German textbooks,
2. to analyze the judges' perceptions in terms of their severity or leniency,
3. to analyze the difficulty of items used in the questionnaire to evaluate high school German textbooks
4. to analyze any bias of judges

2. Method

The Rash model explains how a person's performance with regard to a specific trait predicting that person's response (e.g. right/wrong) and provides valuable data for the development, modification, and monitoring of valid measurement (Boone & Scantlebury, 2006). In this context, a survey research design was employed in the study. Survey research is used to provide a "snapshot of how things are at a specific time" (Denscombe, 1998).

2.1 Study Group

According to the development level of provinces, located in seven regions, as *developed, moderately developed*, and *least developed* (Table 1), a total of 21 teacher trainers, three from each region and selected randomly, were contacted via email and invited to participate in the current study during the 2014–2015 academic year.

Table1. Distribution of the study group according to regions

Regions	Provinces & Teacher Trainers' Numbers	Total
Mediterranean	Adana1, Antalya1, Isparta1, K.Maraş1, Mersin1,	5
Black sea	Ordu1, Samsun1, Trabzon1, Zonguldak1	4
Aegean	Aydın1, Denizli1, İzmir2,	4
Marmara	Balıkesir1, İstanbul3, Sakarya1, Yalova1	6
Central Anatolia	Ankara4, Çankırı1, Kayseri1, Kırşehir1, Niğde1, Sivas1,	9
Eastern Anatolia	Elazığ1, Erzurum2, Malatya2, Tunceli2,	7
Southeastern Anatolia	Diyarbakır2, Urfa1,	3
7 Regions	28 Provinces	38

According to the development level of these provinces as developed*, moderately developed,* and *least developed, a* total of 21 teacher trainers, three selected randomly from each region, were contacted via email and participated in the current study.

Materials used in the study are;

A1.1 Deutschstube (İncebel, Balkan, & Dülger, 2014)

A1.2 Deutsch Training A1 (Kalkan & Çiftarslan, 2013)

A2.1 Hallo Kursbuch & Übungsbuch (Başarmış, 2014)

A2.2 Deutsch Training A2 (Kalkan & Çiftarslan, 2013)

2.2 Data Collection

In the study, a questionnaire related to High School German Textbook Evaluation was prepared to collect the quantitative data (Appendix-1) in the light of review of literature and experts (2 Associate Professors of German, 1 Assistant Professor of German, 3 German lecturers, of two have master's degree, 1 Turkish teacher and 3 German teachers). The questionnaire which was developed in line with Batdı's (2010) scale that he had used in his MA thesis included a 5-point Likert type scale with five options, from '*Strongly Disagree'*, to '*Strongly Agree'*.

According to Lawshe (1975), a content validity ratio (CVR) of .56 would be required to retain the item. Content validity index (CVI) value for each item was computed separately. Experts were asked to score the relevance of each item as 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant (e.g., Davis, 1992). Then the CVI, for each item, was computed as the number of experts giving a scoring of either 3 or 4, divided by the number of experts. A CVI of .80 is considered an acceptable value for good content validity (Yurdugül, 2005). The items' content validity indices (CVIs) determined as criterion for content validity ratios, were found to be 0.82. Since this value is larger than the 0.56 content validity criterion (CVC) [(0.82>0.56) (CVI>CVC)], it can be said that the content validity of items in the questionnaire are statistically significant at the 0.05 level (Veneziano & Hooper 1997).

2.3 Data Analysis

In the analysis of data, FACETS analysis program in which the Rasch measurement model described by Linacre (1993) was used. Three facets of the study according to the Rasch measurement model were as follows:

1) Judges, 21 German teacher trainers

2) Items related to high school German textbooks (11 items).

3) German textbooks (A1.1, A1.2, A2.1, A2.2) for 9^{th}, 10^{th}, 11^{th}, and 12^{th}grades

Data calibration map related to the facets is given in Figure1.

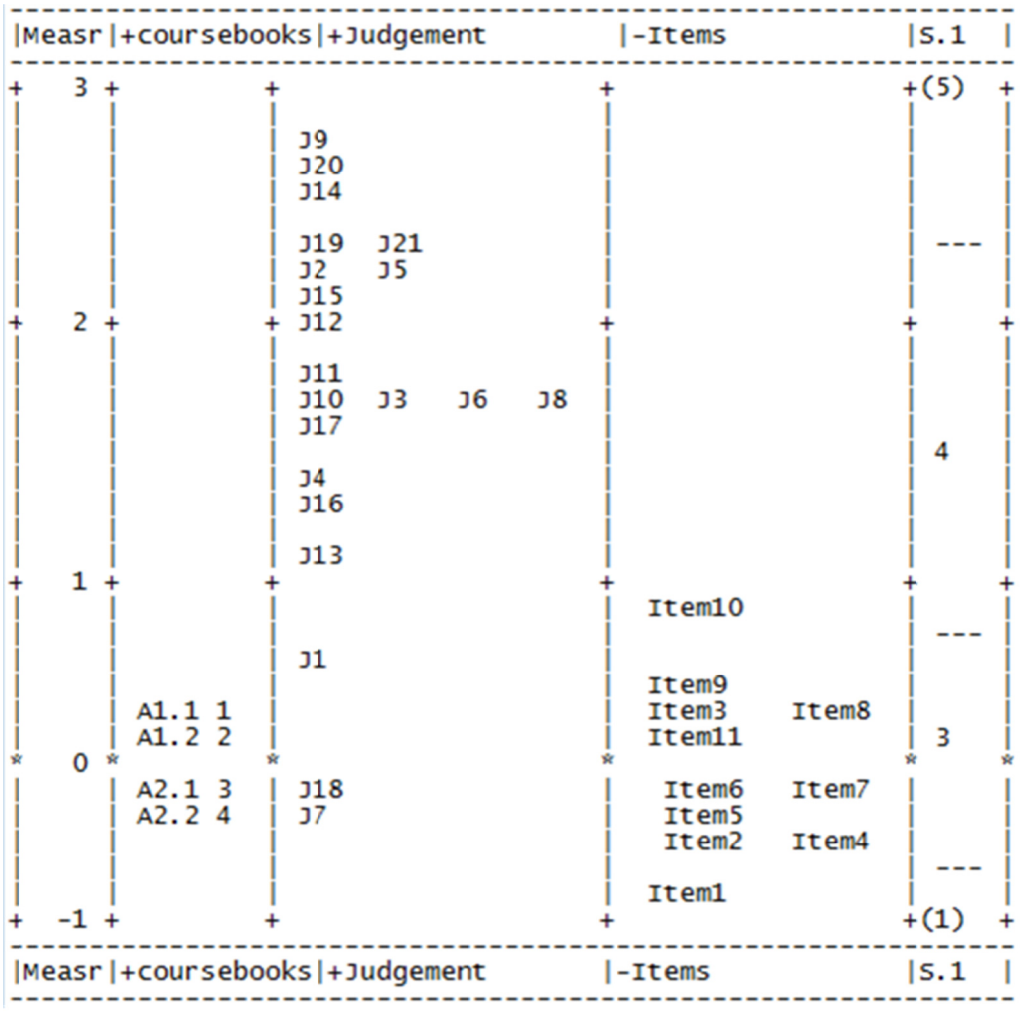

Figure 1. Data Calibration Map

Data obtained from three facets were specified for this study as:

a) A1.1, A1.2, A2.1, A2.2 textbooks for 9th, 10th, 11th, and 12thgrades

b) Leniency/severity of judges,

c) suitability of items,

In Figure1, German textbooks, judges and items were given separately. According to the "German textbooks" column, while the textbook coded A1.1 has the highest quality, the textbook coded A2.2 has the poorest quality. On the other hand, in "judges" column, J(udge)7 has the most severe and J9 has the most lenient behavior. When the column in which items used to evaluate German textbooks is examined, the most difficult item is 10- *Various measuring instruments (matching, short answer test, etc.)are available at the end of units* - while the easiest item is 1- *Objectives are appropriate to grade level*.

3. Results

The views of German teacher trainers who participated in the study on high school German textbooks were analyzed within the framework of evaluation forms via the manyfacet Rasch model which allows for the systematic analysis of coders, judges, or evaluators (Lunz & Linacre, 1998).

3.1 High School German Textbooks

Measurements related to high school German textbooks were presented in Table 2 comprehensively. According to Table 2, while the reliability co-efficient in Rasch analysis is .77 which indicates a high reliability of related textbooks' rankings, the separation index is 1.81. In line with the results, it is said that there are statistically significant differences among German textbooks ($\chi 2$ =17.3, d.f. = 3, p = 0.00). The ranking of German textbooks from the most adequate to

inadequate is as follows: “A1.1, A1.2, A2.1, and A2.2”.

Table 2. High school German textbooks’ measurement report

Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	N coursebooks
965	231	4.2	4.26	.20	.09	1.0	0	1.0	0	1 A1.1 1
959	231	4.2	4.23	.15	.09	1.1	0	1.0	0	2 A1.2 2
927	231	4.0	4.08	-.10	.09	1.0	0	0.9	0	3 A2.1 3
906	231	3.9	3.97	-.25	.08	1.0	0	1.0	0	4 A2.2 4
939.3	231.0	4.1	4.14	.00	.09	1.0	0.0	1.0	-0.2	Mean (Count: 4)
24.0	0.0	0.1	0.12	.18	.00	0.0	0.4	0.0	0.4	S.D.

RMSE (Model) .09 Adj S.D. .16 Separation 1.81 Reliability .77
Fixed (all same) chi-square: 17.3 d.f.: 3 significance: .00
Random (normal) chi-square: 3.0 d.f.: 2 significance: .22

“Infit” and “outfit” statistical values related to the facets in Rasch analysis are also given in Table 2. The quality control limit for both values is between the range of 0.6–1.4 (Wright & Linacre, 1994). Although the infit index is a value showing sensibility to unexpected answers at the point of decision-making, the outfit index is a value showing sensibility to unexpected answers which are outlier (Bastürk, 2010). According to Table 1, all the values are observed not to exceed the determined limit for both indices (1.5).

3.2 Analysis of Judges

Table 3 presents information about leniency/severity of judges regarding the evaluation of German textbooks. When judges are ranked from the severest to the most lenient, J7 is the most severe and J9 is the most lenient. Except from the values in extreme limits, the standard error (RMSE) related to the judges’ severity/ leniency is the calculated value including all the data error measurements. When this value is at 0.20, it indicates that standard error is quite low. In addition, the adjusted standard deviation value considering the relevant error rate is (0.74), below the critical value of 1.0. The reliability co-efficient related to the judges’ scoring behaviors and calculated as 0.93 indicates that judges’ scoring behaviors have been performed at a high reliability.

Table 3. Judges’ measurement report

Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Nu Judgement
203	44	4.6	4.64	2.70	.26	0.8	0	0.8	0	9 J9
201	44	4.6	4.59	2.57	.25	1.0	0	1.1	0	20 J20
199	44	4.5	4.55	2.45	.24	1.0	0	1.1	0	14 J14
197	44	4.5	4.50	2.34	.23	1.0	0	0.9	0	19 J19
196	44	4.5	4.48	2.29	.22	0.7	-1	0.7	-1	21 J21
194	44	4.4	4.44	2.20	.22	1.0	0	1.0	0	2 J2
194	44	4.4	4.44	2.20	.22	0.9	0	0.9	0	5 J5
191	44	4.3	4.37	2.06	.21	1.2	0	1.1	0	15 J15
190	44	4.3	4.34	2.01	.21	0.9	0	0.9	0	12 J12
184	44	4.2	4.21	1.77	.20	0.8	-1	0.8	-1	11 J11
183	44	4.2	4.18	1.74	.19	1.2	0	1.1	0	3 J3
181	44	4.1	4.14	1.66	.19	1.1	0	1.1	0	6 J6
181	44	4.1	4.14	1.66	.19	1.4	1	1.4	1	8 J8
181	44	4.1	4.14	1.66	.19	1.2	1	1.2	0	10 J10
180	44	4.1	4.11	1.63	.19	0.9	0	0.8	0	17 J17
173	44	3.9	3.95	1.38	.18	1.4	2	1.5	2	4 J4
170	44	3.9	3.88	1.29	.18	0.6	-2	0.6	-2	16 J16
165	44	3.8	3.76	1.13	.18	0.6	-2	0.7	-2	13 J13
146	44	3.3	3.31	.55	.17	1.2	0	1.2	0	1 J1
125	44	2.8	2.83	-.10	.18	0.8	-1	0.8	-1	18 J18
123	44	2.8	2.78	-.16	.18	1.1	0	1.1	0	7 J7
178.9	44.0	4.1	4.08	1.67	.20	1.0	-0.1	1.0	-0.2	Mean (Count: 21)
22.1	0.0	0.5	0.52	.77	.02	0.2	1.2	0.2	1.2	S.D.

RMSE (Model) .20 Adj S.D. .74 Separation 3.62 Reliability .93
Fixed (all same) chi-square: 323.0 d.f.: 20 significance: .00
Random (normal) chi-square: 20.1 d.f.: 19 significance: .39

As shown in Table 3, the Judge Separation Index is 3.62 and the reliability co-efficient is 0.93. When the hypothesis “There are statistically significant differences among judges in terms of the degrees of severity/leniency” is tested by Chi-Square ($\chi 2$= 323.0, d.f.=20, p=0.00), the null hypothesis is rejected. In other words, it is emphasized that judges have shown statistically significant differences among themselves. Furthermore, while the outfit value of J4 falls

outside the range of 0.6 – 1.4 which is the accepted value proposed by Bond and Fox (2007) and Wright and Linacre (1994), the infit and outfit values of the other 20 judges included in the study are within the acceptable range, therefore, suitable. Since the mean square value of outfit belonging to J4 is higher than the expected values, this judge is unlikely to have consistent scoring behaviors in the evaluation of German textbooks.

3.3 The Analysis of Items Used to Evaluate German Textbooks

Table 4 presents information related to whether the items to measure German textbooks fit for purpose or not. While the item 10 was evaluated as the most difficult item, item 1 was found to be the easiest one among the 11 items by the participants.

Table 4. The analysis of items used to evaluate German textbooks

Obsvd Score	Obsvd Count	Obsvd Average	Fair Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Nu	ItemS
292	84	3.5	3.47	.92	.13	1.4	2	1.4	2	10	Item10
330	84	3.9	3.97	.25	.14	1.3	2	1.3	1	9	Item9
332	84	4.0	4.00	.22	.14	1.0	0	1.0	0	8	Item8
334	84	4.0	4.02	.18	.14	0.8	-1	0.8	-1	3	Item3
337	84	4.0	4.06	.12	.14	0.9	0	0.9	0	11	Item11
349	84	4.2	4.22	-.13	.15	0.8	-1	0.8	-1	6	Item6
349	84	4.2	4.22	-.13	.15	0.7	-2	0.7	-2	7	Item7
350	84	4.2	4.23	-.15	.15	0.7	-2	0.7	-1	5	Item5
355	84	4.2	4.30	-.26	.15	1.2	0	1.2	0	2	Item2
355	84	4.2	4.30	-.26	.15	1.1	0	1.0	0	4	Item4
374	84	4.5	4.54	-.75	.17	1.0	0	1.1	0	1	Item1
341.5	84.0	4.1	4.12	.00	.15	1.0	-0.2	1.0	-0.2	Mean (Count: 11)	
19.9	0.0	0.2	0.26	.40	.01	0.2	1.5	0.2	1.5	S.D.	

RMSE (Model) .15 Adj S.D. .37 Separation 2.55 Reliability .87
Fixed (all same) chi-square: 85.1 d.f.: 10 significance: .00
Random (normal) chi-square: 10.0 d.f.: 9 significance: .35

The Standard Error (RMSE), related to the analysis of the items used to evaluate ECM is 0.15, which is the low value in determining the quality. The standard deviation value corrected for estimation error has been calculated as 0.37, which is below the critical value of 1.0. While the Separation Index is 2.55 and the Reliability Co-efficient is 0.87. When the hypothesis "There are statistically significant differences in terms of item difficulties indicating the quality of German textbooks" is tested by Chi-Square ($\chi 2$= 85.1, d.f.=10, p=0.00), the null hypothesis is rejected. In other words, it is emphasized that items evaluate different characteristics belonging to the textbooks and have shown statistically significant differences.

When "infit" and "outfit" values related to facets are examined, all the textbooks have acceptable level of 1.5. This result indicates that almost all items are consistent with the evaluation of the textbooks and their infit and outfit mean squares are within the acceptable values.

3.4 Judges' Bias Interaction Analysis

In Table 5, interaction analysis related to the views of the judges on German textbooks is presented. According to Semerci (2011), Z points lying outside +2 and -2 are signs of interaction bias. In table 5, Z points vary between 2.90 and -2.22, indicating that judges made extremely severe or lenient evaluations on German textbooks. In this context, J8 gave 39 points (Z=2.90) and exhibited severe bias to the textbook coded A1.1, but should have given 47 points. Similarly, instead of 44 points, J6 gave 37 points (Z=2.32) for the textbook coded A2.2 and exhibited severe bias.

Table 5. Interaction analysis of high school German textbooks evaluated by judges

Obsvd Score	Exp. Score	Obsvd Count	Obs-Exp Average	Bias+ Model Measure	Model S.E.	Z-Score	Infit MnSq	Outfit MnSq	Sq	N	c.books	measr	Nu	Puan measr
39	46.6	11	-.69	1.01	.35	2.90	0.9	1.0	29	1	A1.1 1	.20	8 P8	1.66
37	43.5	11	-.59	.80	.35	2.32	1.1	1.1	24	4	A2.2 4	-.25	6 P6	1.66
54	51.3	11	.24	-1.37	1.01	-1.36	0.9	0.8	34	2	A1.2 2	.15	9 P9	2.70
54	51.0	11	.27	-1.45	1.01	-1.44	0.9	0.8	77	1	A1.1 1	.20	20 P20	2.57
54	49.9	11	.37	-1.73	1.01	-1.71	0.9	0.8	81	1	A1.1 1	.20	21 P21	2.29
51	44.7	11	.57	-1.17	.53	-2.22	1.8	1.2	13	1	A1.1 1	.20	4 P4	1.38
44.7	44.7	11.0	.00	-.07	.43	-.01	1.0	0.9	Mean (Count: 84)					
6.2	5.7	0.0	.25	.50	.13	1.04	1.0	0.3	S.D.					

Fixed (all = 0) chi-square: 91.3 d.f.: 84 significance: .27

Besides severe biases, lenient behaviors are also exhibited by the judges included in the study. For example, J9 gave 54 points (Z= -1.36) instead of 51 points for the textbook coded A1.2 and exhibited lenient behaviors. Similarly, J20 gave

54 points (Z=-1.44) instead of 51 points for the textbook coded A1.1; J21 gave 54 points (Z= -1.71) instead of 50 points for the textbook coded A.1.1; J4 gave 51 points (Z= -2.22); but should have given 45 points for the textbook coded A1.1 and they exhibited extremely lenient behaviors.

4. Discussion

In this study, data related to high school German textbooks were analyzed using the many-facets Rasch measurement model. Three facets were specified as German textbooks, Leniency/severity of judges, and suitability of items. According to results, while the textbook coded A1.1 had the highest quality, the textbook coded A2.2had the poorest quality among the German textbooks for 9^{th}, 10^{th}, 11^{th}, and 12^{th}grades. In terms of Judges' bias on the evaluation of high school German textbooks, while J8 (for A1.1) and J6 (for A2.2) exhibited severe bias, J9 (for A1.2); J20 (for A1.1); J21 (for A1.1); and J4 (for A1.1) exhibited extremely lenient behaviors. According to the items prepared to evaluate "high school German textbooks", the most difficult item was 10 "Various measuring instruments (matching, short answer test, etc.) are available at the end of units "while the easiest item was 1"Objectives are appropriate to grade level". For judges, J7 has the most severe; J9 has the most lenient behavior. In terms of "infit" and "outfit" values, except for J4 whose outfit value exceeded the limit value, all the other 20 judges included in the study were within the acceptable range (0.6- 1.4), therefore, suitable. Since the mean square value of outfit belonging to J4is higher than the expected values, this judge is unlikely to have consistent scoring behaviors in the evaluation of German textbooks. In other words, there was a statistically significant difference between leniency and severity of judges. Similarly, Batdı (2013, 2014) found statistically significant differences between leniency and severity of judges in his studies that are related to the evaluation of high school English and Maths curriculum respectively. According to Bastürk (2010), the Rasch measurement model gives a better reliability result which is similar to Cronbach's alpha reliability co-efficient. As interpreted in the traditional reliability results, the closer the correlation comes to +1.00, the more reliable the test is (Bastürk, 2010, p. 57). In the current study, a reliability value of 1.00 for determination of the quality of German textbooks, 0.99 for determination of the judges' severity/leniency levels and 0.89 for determination of difficulty or easiness of the items has been obtained. In the light of the results, more rigorous and detailed studies are suggested to improve the quality of textbooks. Since some German teacher trainers displayed biased behaviors as judges both positively and negatively, it is also suggested that teachers should be unbiased when evaluating everything for students.

References

Allwright, R. (1982). What do we want teaching materials for? *ELT Journal*, *36*(1), 131-147. http://dx.doi.org/10.1093/elt/36.1.5

Başarmış, Y. (2014). *Hallo Kursbuch & Übungsbuch*. Ankara: Almanca Yayınlar. Genel Almanca.

Baştürk, R. (2010). Bilimsel araştırma ödevlerinin çok yüzeyli Rasch ölçme modeli ile değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *1*(1), 51-57. ISSN: 1309-6575.

Batdı, V. (2010). *Yabancı dil ağırlıklı liselerde okutulan ingilizce ders kitabının etkili kullanılabilmesi için taşıması gereken özelliklere ilişkin öğretmen görüşlerinin değerlendirilmesi* (Elazığ ve Diyarbakır illeri örneği). [*An assessment of teachers'opinions about the effectiveness of English coursebook used in foreign language high schools (case of Elazig and Diyarbakır cities*)]. (Unpublished master's thesis). University of Firat, Elazig, Turkey.

Batdı, V. (2013). Evaluation of teachers' views about the affective domain of the high school new English curriculum in Turkey. *Mediterranean Journal of Educational Research, 14a*, 221-28.

Batdı, V**.** (2014). Ortaöğretim matematik öğretim programı içeriğinin Rasch ölçme modeli ve Nvıvo ile analizi. *Turkish Studies - International Periodical for the Languages, Literature and History of Turkish or Turkic Volume, 9*(11), 93-109. http://dx.doi.org/10.7827/turkishstudies.7074

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences*. (2nd Edition). New Jersey: Lawrence Erlbaum Inc. Publishers

Boone, W., & Scantlebury; K. (2006). The role of Rash analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*, 253-269. http://dx.doi.org/10.1002/sce.20106

Crawford, J. (2002). The role of materials in the language classroom: Finding the balance. In J. C. Richards, & W. A. Renandaya (Eds.), *Methodology in language teaching: An anthology of current practices* (pp.80-95). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667190.013

Crystal, D. (1987). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.

Cunningsworth, A. (1995). *Choosing your coursebook*, Oxford: Heinemann.

Chatman, S. P., & Goetz, E. T. (1985). Improving textbook selection. *Teaching of Psychology*, *12*, 150-152.

http://dx.doi.org/10.1207/s15328023top1203_9

Dar, F. (2012). Textbook materials and their successful application in the classroom: Implications for language development. *Journal of Educational and Instructional Studies in the World, 2*(4), 109-114. ISSN: 2146-7463

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, 5, 194-197. http://dx.doi.org/10.1016/S0897-1897(05)80008-4

Denscombe, M. (1998). *The good research guide: For small-scale social research projects*. Buckingham: Open University Press.

Dudley-Evans, T., & St. John, M. J. (1998). *Developments in English for specific purposes: A multidisciplinary approach.* Cambridge: Cambridge University Press.

Garinger, D. (2002). Textbook selection for the ESL classroom. Retrieved from http://www.cal.org/resources/digest/0210garinger.html

Gray J. (2002). The global coursebook in English language teaching. In Block D., Cameron D. (Eds.), *Globalization and language teaching* (pp. 151-167). London, England: Routledge.

İncebel, B., Balkan, F., & Dülger, M. (2014). *Deutschstube* (1st Ed.). Ankara: Milli Eğitim Bakanlığı Yayınları.

Kalkan, K., & Çiftarslan,S. (2013). *Deutsch Training*. Ankara: Almanca Yayınlar. Genel Almanca

Kitao, K. S., & Kitao, K. (1997) Selecting and developing teaching/Learning materials. *The Internet TESL Journal, IV*, 4.

Küçükahmet, L. (1995). *Öğretim ilke ve yöntemleri* (7th ed.). Ankara: Gazi Büro Kitabevi, Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*, 563-575. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.9380&rep=rep1&type=pdf

Linacre, J. M. (1993). *Generalizability theory and many-facet Rasch measurement*. A paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.(April 12-16, 1993). Retrieved from http://files.eric.ed.gov/fulltext/ED364573.pdf

Lunz, M. E., & Linacre, J. M. (1998). Measurement designs using multifacet Rasch modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research. Methodology for business and management* (pp. 47–77). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mohammadi, M., & Abdi, H. (2014). Textbook evaluation: A case study. *Procedia-Social and Behavioral Sciences*, *98*, 1148-1155. http://dx.doi.org/10.1016/j.sbspro.2014.03.528

Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. New York: Routledge.

NCTE, Executive Committee. (2014). Guidelines for selection of materials in English Language Arts Programs.

Retrieved fromhttp://www.ncte.org/positions/statements/material-selection-ela

Parrish, B. (2004). *Teaching adult ESL: A practical introduction*. McGraw Hill: New York, NY.

Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667220

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667190

Semerci, Ç. (2011): Mikro öğretim uygulamalarının çok yüzeyli Rasch ölçme modeli ile analizi. *Eğitim ve Bilim*, *36*(161), 14-25. Retrieved from http://egitimvebilim.ted.org.tr/index.php/EB/article/view/145/281

Sheikhzadeh, M. E. (2011). Adoption, adaptation, and development of language instructional units. *European Journal of Social Sciences*, *22*(4), 550-555. Retrieved from http://arzek.weebly.com/uploads/1/2/9/1/12914289/adopt_n_adapt.pdf

Sikorova, Z. (2011). The role of textbooks in lower secondary schools in the Czech Republic. *IARTEM e-Journal*, *4*(2), 1-22.Retrieved from http://biriwa.com/iartem/ejournal/volume4.2/papers/Paper1_Sikorova_IARTEM_RoleTextbook_eJournal_Vol4No2.pdf

Tomlinson, B. (2001). Materials development. In R. Carter, and D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 66-71). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667206.010

Veneziano, L., & Hooper, J. A. (1997). Method for quantifying content validity of health-related questionnaires *American Journal of Health Behavior*, *21*(1), 67-70.

Wen, C. W., Chien, H. L., & Chung, C. L. (2011). Thinking of the Textbook in the ESL/EFL classroom. *English Language Teaching*, *4*(2), 91-96. http://dx.doi.org/10.5539/elt.v4n2p91

Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal*, *37*(3), 251–255. http://dx.doi.org/10.1093/elt/37.3.251

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. Transactions of the Rasch measurement SIG American educational research association. *Rasch Measurement Transaction, 8*(3), 370-382. Retrieved from http://www.rasch.org/rmt/rmt83b.htm

Yurdugül, H. (2005). *Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması*. Paper presented at the XIV. National Educational Sciences Congress,Denizli, Turkey (28-30 September 2005). Retrieved from http://yunus.hacettepe.edu.tr/~yurdugul/3/indir/PamukkaleBildiri.pdf

Appendix. A Questionnaire Belonging to the Evaluation of High School German Textbooks

Dear Colleague;

The aim of this study is to determine German **teachers' views on the German Textbooks**. Please select the appropriate option for each item by specifying in numbers as: "**1: Totally Agree 2: Mostly Agree 3: Partly Agree 4: Often Disagree 5: Disagree**". We thank you for your help and wish you success in your professional life

ITEM NUMBER

1. Gender? □ Male □ Female

2. The city where you work: ……..

3. School type that you work?
□ Science High School □ Anatolian Teacher High School
□ Anatolian High School □Technical-Vocational High School
□ Regular High School □ Other:……....

4. Seniority years of service
□ 1–5 years □ 6-10 years □ 11-15 years □16-20years □ 21+ years

5. Faculty / Department you graduated
□ Education Faculty □ Faculty of Literature
□ Other:……..

	German textbooks for 9th, 10th, 11th and 12th grades	**A1.1**	**A1.2**	**A2.1**	**A2.2**
1	Objectives are appropriate to grade level.				
2	Objectives are associated with the content. .				
3	Objectives are consistent with the assessments found end of units .				
4	Content is valid and reliable.				
5	Content is suitable for teaching principles.				
6	Visual elements of content are sufficient.				
7	Activities develop the critical thinking skills.				
8	The textbook is intended to develop basic language skills (Reading, Writing, Speaking, Listening)				
9	Expression is provided by modern methods and techniques (Eclectic methods, etc.) .				
10	Various measuring instruments (matching, short answer test, etc.)are available at the end of units				
11	Questions are appropriate for cognitive taxonomy (recall, comprehension, application, analysis, comparison and creation.)				